Sampling Potential Energy Surface of Glycyl Glycine Peptide: Comparison of Metropolis Monte Carlo and Stochastic Dynamics

HANOCH SENDEROWITZ, W. CLARK STILL
Department of Chemistry, Columbia University, New York, New York 10027

ABSTRACT: A comparative study was carried out to test the efficiency with which Metropolis Monte Carlo (MC) and stochastic dynamics (SD) sample the potential energy surface of the *N*-acetyl glycyl glycine methylamide peptide as defined by the united atom AMBER* force field. Boltzmann-weighted ensembles were generated with variations of all internal degrees of freedom (i.e., stretch, bend, and torsion) for a single *N*-acetyl glycyl glycine methylamide molecule at 300 K by 10^8-step MC and 100-ns SD simulations. As expected, both methods gave the same final energetic results. However, convergence was found to be ~10 times faster with MC than with SD as measured by comparisons of the populations of all symmetrically equivalent conformers. © 1998 John Wiley & Sons, Inc. J Comput Chem 19: 1294–1299, 1998

Keywords: Monte Carlo; molecular dynamics; free energy simulation; simulation convergence

Introduction

The two most common methods for modeling organic and biological molecules are Monte Carlo (MC) and molecular or stochastic dynamics (MD or SD).[1] Both rely on classical potential energy functions to calculate molecular energies and, in the case of dynamics, derivatives. The main difference is in the method of sampling the potential energy surface available to the system. For MC, new molecular configurations are typically generated by making random structural changes to the molecule's internal degrees of freedom and then accepting or rejecting those configurations based on the Metropolis test.[2] Using this procedure to generate a sufficiently large number of configurations leads to a Boltzmann-weighted ensemble. For MD (or SD), new configurations are obtained by applying Newton's equations of motion(Langevin's equations for SD) to update

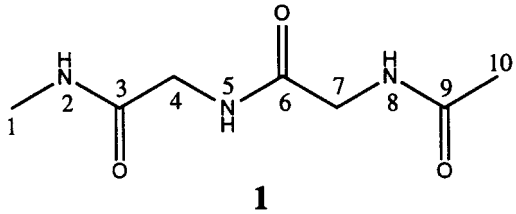
Correspondence to: H. Senderowitz; e-mail: hanoch@ala.deptor.co.il

CCC 0192-8651 / 98 / 111294-06

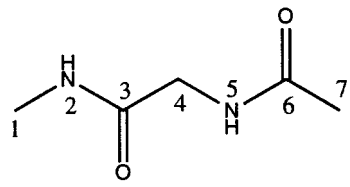
atomic positions and velocities repetitively. The resulting trajectory explores the system's potential energy surface and leads ultimately to the same ensemble of states that is produced by MC in the long simulation limit.

Considering the importance and the many applications of those two simulation methods, surprisingly few examples are found in the literature in which their relative efficiencies were thoroughly investigated. Thus, Northrup and McCammon found MC to be 10 times less efficient than MD in sampling the potential energy surface of bovine pancreatic trypsin inhibitor (BPTI) in the gas phase.[3] However, these authors used a rather inefficient MC procedure whereby new configurations were generated by single atomic movements in Cartesian space. MC methods based on structural changes to the molecule's internal degrees of freedom (i.e., stretch, bend, torsion) might be expected to lead to a much improved MC sampling of the potential energy surface (see below). Thus, Guarnieri and Still found internal coordinate MC to converge at least 10 times faster than SD with united atom (5 particle) *n*-pentane at 300 K.[4] This conclusion was supported in a recent study by Jorgensen and Tirado–Rives who found that MD calculations required 1.6–3.8 times more computer time to achieve the same level of convergence as MC simulations for liquid *n*-hexane.[5]

It is well recognized that free energy based molecular properties can be obtained only with adequate sampling of the entire potential energy surface, so simulation convergence is crucial to successful molecular modeling.[6] To test the relative efficiencies of MC and SD with a small biological molecule, we carried out lengthy simulations using both methods on a single *N*-acetyl glycyl glycine methylamide molecule (**1**) in the gas phase. This system is attractive because it embodies many key features of peptidic chains and has torsional barriers that are typical of bioorganic molecules. Furthermore, some pairs of **1**'s conformers are enantiomeric (i.e., they are mirror image isomers and thus energeticaly equivalent) and therefore provide a convenient method for evaluating simulation convergence.



1



2

Computational Details

All simulations were carried out on a single *N*-acetyl glycyl glycine methylamide molecule (**1**) in the gas phase using the BatchMin V5.5 simulation program as implemented in the MacroModel molecular modeling package.[7] The united atom AMBER* force field was used,[8] and the simulations varied all internal degrees of freedom (i.e., all bond lengths, bond angles, and torsional angles). The nonbonded distance cutoff employed was larger than the maximum dimension of the system.

A conformational search using the internal coordinate SUMM method[9] identified 13 distinct minima on the potential energy surface of **1**. The global minimum had the totally extended conformation (i.e., all backbone torsions = 180°) while the remaining 12 minima were divided into six enantiomeric pairs. These minima were supplied to the simulation program as a means of monitoring conformational populations and evaluating simulation convergence. Conformational populations were determined by comparing the simulated structure to each of the input conformations using a least squares superimposition in Cartesian coordinates. The input conformer having the smallest root mean square deviation from the simulated structure was taken to define it. All simulations were run for a total of 10^8 steps (MC) and 100 ns (SD, 10^8 1-fs steps) on a 150-MHz DEC Alpha 3000/600 workstation.

MC DETAILS

As mentioned above, an efficient MC algorithm involves change to the molecule's internal coordinates rather than to its Cartesian coordinates. For this purpose, BatchMin uses an MC algorithm that is based on a *Z*-matrix representation of **1**. In our simulations the *Z*-matrix bond lengths were var-

ied by ±0.0–0.1 Å, bond angles by ±0–5°, and torsional angles by ±0–60°. Each MC step consisted of randomly choosing two internal coordinates, varying them within their designated ranges, rebuilding the structure from the modified coordinates, and then accepting or rejecting it based on the Metropolis test.[2] This simulation protocol led to an acceptance rate of 33%. The MC simulations reported here were run for 10^8 steps at 300 K.

SD DETAILS

A stochastic velocity Verlet algorithm is used in BatchMin to integrate the Langevin equations of motion.[4] The simulations reported here were run for 100 ns using a time step of 1 fs at 300 K [SD frictional coefficient (γ) = 2.5 ps^{-1}]. In order to sample all degrees of freedom (and thus to be comparable with the above MC simualtions), bond lengths were unconstrained. The average temperature at the end of our 100-ns SD simulation was 299.2 K.

Results and Discussion

The barriers for conformational interconversions in the *N*-acetyl glycyl glycine methylamide peptide (**1**) were estimated from the Ramachandran plot of the closely related *N*-acetyl glycine methylamide (**2**).

This plot was obtained at 5° resolution along both torsional axes and is shown with 1 kcal/mol contour intervals in Figure 1. The global energy minimum (ψ, ϕ = 180°, 180°) is separated from the two enantiomeric minima at ψ = 306.1°, ϕ = 77.2° and ψ = 53.9°, ϕ = 282.8° by energy barriers of ~3 kcal/mol. Energy barriers of such heights would be expected to be crossed by a real molecule on average once every 0.02 ns at 300 K.[10]

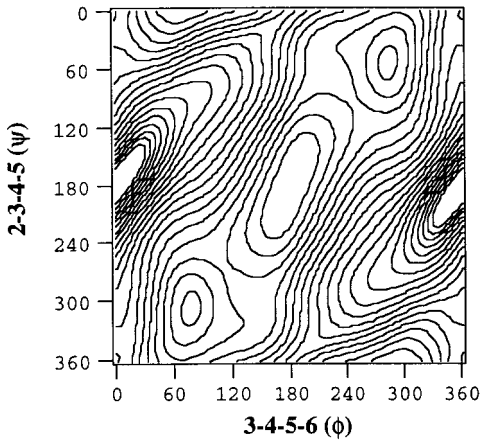


FIGURE 1. Ramachandran plot of the *N*-acetyl glycine methylamide **2** obtained at 5° resolution along both torsional axes. Contours are shown at 1 kcal / mol intervals.

Assuming that **1** and **2** have comparable barriers separating their conformers, quite lengthy simulations of **1** would be needed to assume complete simulation convergence with methods based on MD. Given the 13 distinct minima of **1** and the possibility of some barriers being greater than 3 kcal/mol, we were worried that even the lengthy 100-ns SD simulations we planned might not be sufficient to give stable, converged results. On the other hand, we thought that MC might have a better change of achieving adequate convergence because it might be more efficient at interconverting conformations. In particular, for systems like **1** where conformational space is densely populated, an MC simulation should be able to cross energy barriers efficiently if large, potentially barrier-hopping moves in torsion angle space are including in the simulation.

To judge the ensembles generated by MC and SD we evaluated the first four moments of the potential energy distribution (average, standard deviation, skew and kurtosis)[11] and the populations of **1**'s 13 distinct conformations. These data are summarized in Tables I and II and conform that both MC and SD generate ensembles in the

TABLE I.
Comparison of Ensemble Average Energy (⟨E⟩, kcal / mol), Standard Deviation (SD), Skew (SK), Kurtosis (KU), and Minima Populations between MC (10^8 Steps) and SD (100 ns) Simulations for *N*-Acetyl Glycyl Glycine Methylamide (1) at 300 K *In Vacuo*.

	SD	MC
⟨E⟩	−46.54	−46.54
SD	2.89	2.89
SK	0.10	0.10
KU	0.058	0.060
pop1	0.832	0.824
Ave. pop2–pop3	0.0083	0.0099
Ave. pop4–pop5	0.0094	0.0101
Ave. pop6–pop7	0.0380	0.0378
Ave. pop8–pop9	0.0071	0.0059
Ave. pop10–pop11	0.0030	0.0040
Ave. pop12–pop13	0.0180	0.0203

The minima populations are the actual population of conformer 1 and average populations for conformational pairs 2–3, 4–5, 6–7, 8–9, 10–11, 12–13.

TABLE II.
Comparison of Populations of Six Symmetrically Equivalent Conformational Pairs and Percentage Difference between Them (% diff.). between MC (10^8 steps) and SD (100 ns) Simulations for United Atom *N*-Acetyl Glycyl Glycine Methylamide (1) at 300 K *In Vacuo*.

	SD			MC		
	Population		% Diff.[a]	Population		% Diff.[a]
POP2 −3	0.00658	0.01006	10.5	0.00950	0.01031	2.0
POP4 −5	0.01314	0.00563	20.0	0.01017	0.01008	0.2
POP6 −7	0.04076	0.03547	3.5	0.03768	0.03796	0.2
POP8 −9	0.00496	0.00926	15.1	0.00627	0.00552	3.0
POP10 −11	0.00230	0.00377	12.1	0.00397	0.00400	0.2
POP12 −13	0.01776	0.01810	0.5	0.02095	0.01976	1.5
Average			10.3			1.2

The last entry gives the average percentage difference for the six comparisons.
[a]Calculated from % difference = 100 $|pop_i - pop_j| / 2(pop_i + pop_j)$.

limit of long simulations that are virtually indistinguishable. We attribute the small differences in energy moments and conformational populations that we did observe to the slightly different simulation temperatures (MC, 300 K; SD 299.2 K) and to incomplete convergence of the 100-ns SD run.

In order to assess convergence, we compared the populations of equivalent (enantiomeric) conformational pairs (conformers 2−3, 4−5, 6−7, 8−9, 10−11, 12−13). These are given in Table II along with the percentage difference between the members of each pair calculated from

$$\% \text{ difference} = 100\ |pop_i - pop_j|/2(pop_i + pop_j)$$

where i and j refer to the first and second member of each pair. For the 10^8-step MC simulations, the percentage difference in populations of all enantiomeric conformational pairs at the end of the simulation was found to be equal or less than 3%. For SD, the percentage difference for most conformational pairs at the end of the 100-ns (10^8 SD steps) simulation was above 10% with an average of 10.3%. Thus, for a given length of simulation, the degree of convergence achieved by MC far exceeds that of SD with **1**.

To further quantify the confergence of both simulations, we looked at the time evolution of the conformational populations at constant intervals (MC, 10^6 steps; SD, 1 ns) during the simulations. The results are presented in Figures 2−5. The efficiency of the methods was defined as the time required for the percentage difference between the populations of enantiomeric (thus equally populated at convergence) conformations to fall below some arbitrarily chosen level. For MC, all conformational pairs met this criterion for the 30% level after 4×10^6 steps (64 CPU min), for the 20% level after 6×10^6 steps (97 CPU min), and for the 10% level after 10×10^6 steps (164 CPU min). The corresponding numbers for SD were 17 ns (435 CPU min) for the 30% level and 50 ns (1277 CPU min) for the 20% level. Most conformational pairs did not reach the 10% level of convergence even after 100 ns of SD simulation time. Based on these comparisons, we find MC to be approximately 10 times faster in CPU time than SD in reaching a given degree of convergence with **1** at 300 K.

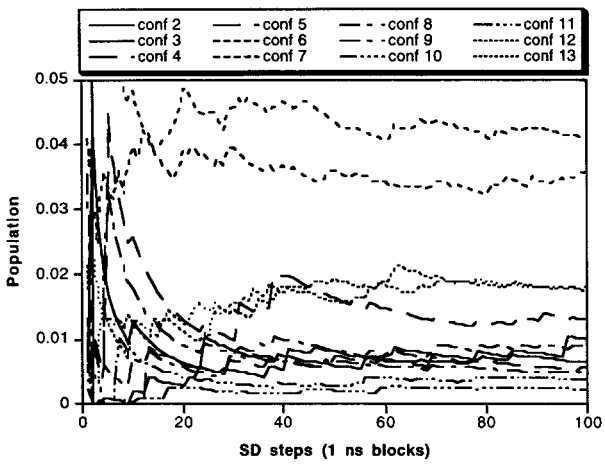


FIGURE 2. Time versus convergence as given by the populations of the six symmetrically equivalent conformational pairs from a 100-ns SD simulation of **1**.

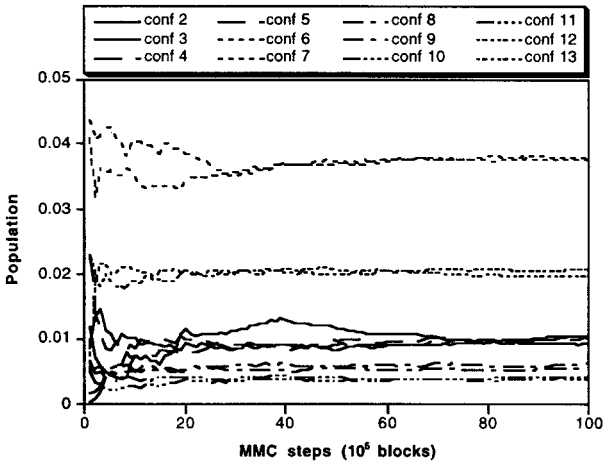


FIGURE 3. Time versus convergence as given by the populations of the six symmetrically equivalent conformational pairs from a 10°-step MC simulation of **1**.

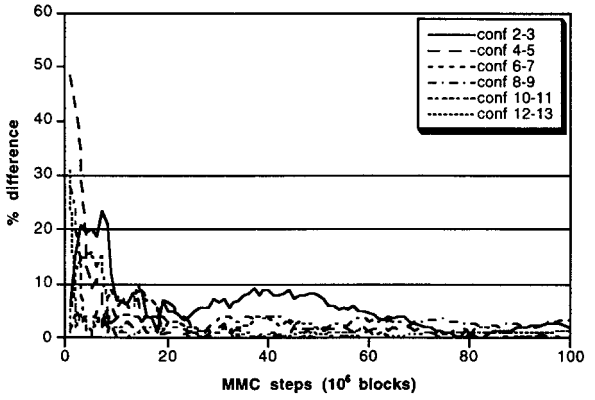


FIGURE 5. Time versus convergence as given by the percentage difference between the populations of the six symmetrically equivalent conformational pairs from a 10^8-step MC simulation of **1**. The three horizontal lines indicate the 30, 20, and 10% levels (see text).

Conclusions

In this work we compared the performance of MC and SD in sampling the potential energy surface of the glycyl glycine peptide **1**. Both simulations were run with the same computer program (BatchMin V5.5) using identical energetic protocols. Although no special attempts were made to optimize the performance of either algorithm for simulations of **1**, we believe that the current BatchMin implementation of both is reasonably efficient.

As we expected, MC and SD both generated emsembles that were indistinguishable in terms of the system's average energy, higher energetic moments, and conformational populations. A few other such comparisons for organic systems have been reported.[5] In terms of simulation efficiency, however, the two methods differed considerably; MC was approximately 10 times faster than SD in reaching a given degree of simulation convergence with **1** at 300 K.

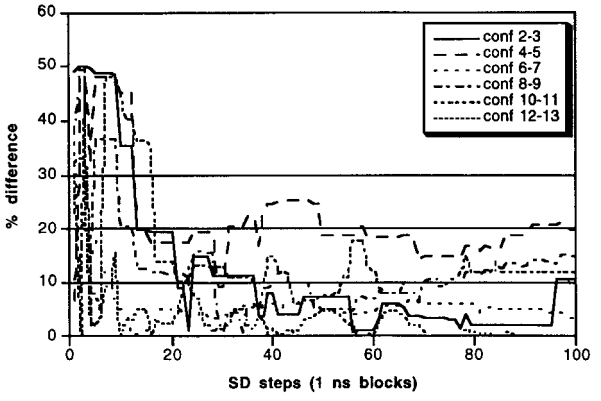


FIGURE 4. Time versus convergence as given by the percentage difference between the populations of the six symmetrically equivalent conformational pairs from a 100-ns SD simulation of **1**. The three horizontal lines indicate the 30, 20, and 10% levels (see text).

N-Acetyl glycyl glycine methylamide is perhaps particularly suited for MC sampling because much of its potential energy surface is densely populated with low energy conformational states. Consequently, large random steps in torsion angle space have a good change of yielding reasonably low energy structures and this feature facilitates conformational interconversions. Obviously, other test systems might present a different picture of relative efficiencies of SD and MC. For example, if the conformational space of a system is only sparsely populated (i.e., when there are a small number of low energy regions on a complex potential energy surface), large MC steps would be more likely to lead to high energy regions of the surface and would therefore result in poor acceptance rates and slow convergence. In such cases the relative efficiencies of MC and SD might be different from what we found here. Nevertheless, the current study along with previous work in this field[4,5] clearly negates an notion of comparative inefficient sampling with MC simulations in flexible, multiconformational systems. Furthermore, problems such as those described here may be largely overcome by modifying the basic Metropolis algorithm into smart MC algorithms, resuting in significantly

improved sampling of the potential energy surface. Several such methods have been described in the literature and are usually referred to as importance sampling methods.[12] Some were shown to be highly efficient for the conformational sampling of multiconformational organic molecules.[13] Thus, MC deserves strong consideration as an efficient simulation method for exploring the potential energy surfaces of flexible molecules.

References

1. M. P. Allen and D. J. Tildesley, *Computer Simulations of Liquids*, Clarendon Press, Oxford, U.K., 1987.
2. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. Teller, and E. Teller, *J. Chem. Phys.*, **21**, 1087 (1953).
3. S. A. Northrup and J. A. McCammon, *Biopolymers*, **19**, 1001 (1980).
4. F. Guarnieri and W. C. Still, *J. Comput. Chem.*, **15**, 1302 (1994).
5. W. L. Jorgensen and J. Tirado–Rives, *J. Phys. Chem.* **100**, 14508 (1996).
6. (a) W. L. Jorgensen, *J. Phys. Chem.*, **87**, 5304 (1983); (b) C. L. Brooks III, B. M. Pettitt, and M. Karplus, *Adv. Chem. Phys.*, **71**, 1 (1988); (c) P. A. Kollman, *Chem Rev.*, **93**, 2395 (1993); (d) M. J. Mitchell and J. A. McCammon, *J. Comput. Chem.* **12**, 271 (1991); (e) W. F. van Gunsteren and A. E. Mark, *Eur. J. Biochem.*, **204**, 947 (1992); (f) A. E. Mark, S. P. van Helden, P. E. Smith, L. H. M. Janssen, and W. F. van Gunsteren, *J. Am. Chem. Soc.*, **116**, 6293 (1994).
7. F. Mohamadi, N. G. J. Richards, W. C. Guida, R. Liskamp, M. Lipton, C. Caufield, G. Chang, T. Hendrickson, W. C. Still, *J. Comput. Chem.*, **11**, 440 (1990).
8. S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, and P. Weiner, *J. Am. Chem. Soc.*, **106**, 765 (1984).
9. J. M. Goodman and W. C. Still, *J. Comput. Chem.*, **12**, 1110 (1991).
10. E. L. Eliel and S. H. Wilen, *Stereochemistry of Organic Compounds*, Wiley, New York, 1994, p. 635.
11. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes*, Cambridge University Press, New York, 1986, p. 455.
12. (a) M. Rao, C. Pangali, and B. Berne, *J. Mol. Phys.* **37**, 1773 (1979); (b) C. Pangali, M. Rao, B. Berne, *J. Chem. Phys. Lett.* **55**, 413 (1978); (c) M. Rao and B. J. Gerne, *J. Chem. Phys.*, **71**, 129 (1979); (d) J. Cao and B. J. Berne, *J. Chem. Phys.* **92**, 1980 (1990); (e) P. J. Rossky, J. D. Doll, and H. L. Friedman, *J. Chem. Phys.*, **69**, 4628 (1978); (f) S. Goldman, *J. Chem. Phys.*, **79**, 3938 (1983); (g) M. Mezei, K. A. Bencsath, S. Goldman, and S. Singh, *Mol. Simul.*, **1**, 87 (1987); (h) M. J. Kotelyanskii and U. W. Suter, *J. Chem. Phys.*, **96**, 5383 (1992); (i) T. Noguti and N. Go, *Biopolymers*, **24**, 527 (1985); (j) R. Abagyan and M. Totrov, *J. Mol. Biol.* **235**, 983 (1994); (k) M. E. Clamp, P. G. Baker, C. J. Stirling, and A. Brass, *J. Comput. Chem.* **15**, 838 (1994).
13. (a) H. Senderowitz, F. Guarnieri and W. C. Still, *J. Am. Chem. Soc.*, **117**, 8211 (1995); (b) H. Senderowitz, C. Parish, and W. C. Still, *J. Am. Chem. Soc.* **118**, 2078 (1996); (c) H. Senderowitz and W. C. Still, *J. Phys. Chem. B*, **101**, 1409 (1997); (d) H. Senderowitz and W. C. Still, *J. Org. Chem.*, **62**, 1427 (1997).